DEVELOPMENTAL DYNAMICS 225:384–391 (2002)

REVIEWS

A PEER REVIEWED FORUM

Genetic and Genomic Tools for *Xenopus* Research: The NIH *Xenopus* Initiative

STEVEN L. KLEIN,[1*] ROBERT L. STRAUSBERG,[2] LUKAS WAGNER,[3] JOAN PONTIUS,[3]
SANDRA W. CLIFTON,[4] PAUL RICHARDSON,[5]
THE WASHINGTON UNIVERSITY EST SEQUENCING GROUP,[4] AND
THE NATIONAL INSTITUTES OF HEALTH *XENOPUS* WORKING GROUP[6]

[1]*Developmental Biology, Genetics and Teratology Branch, National Institute of Child Health and Human Development, Rockville, Maryland*
[2]*Cancer Genomics Office, National Cancer Institute, Bethesda, Maryland*
[3]*National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland*
[4]*Genome Sequencing Center, Washington University School of Medicine, St. Louis, Missouri*
[5]*Functional Genomics, US Department of Energy Joint Genome Institute, Walnut Creek, California*
[6] *Members listed at http:/ /www.nih.gov/science/models/xenopus/*

ABSTRACT **The NIH *Xenopus* Initiative is establishing many of the genetic and genomic resources that have been recommended by the *Xenopus* research community. These resources include cDNA libraries, expressed sequence tags, full-length cDNA sequences, genomic libraries, pilot projects to mutagenize and phenotype *X. tropicalis*, and sequencing the *X. tropicalis* genome. This review describes the status of these projects and explains how to access their data and resources. Current information about these activities is available on the NIH *Xenopus* Web site (http://www.nih.gov/science/models/xenopus/).**

Key words: *Xenopus laevis*; *Xenopus tropicalis*; resources; ESTs; cDNA libraries; genomic libraries

HISTORY OF THE NIH *XENOPUS* INITIATIVE

The *Xenopus* Initiative of the US National Institutes of Health (NIH) coordinates NIH activities that are developing genetic and genomic resources for *Xenopus* research. The NIH *Xenopus* Initiative has its origins in the NIH Non-Mammalian Models Initiative, which in turn had its genesis in the Human Genome Project. The Human Genome Project officially began in the late 1980s as a joint effort of the US Department of Energy (DoE) and the NIH, and subsequently expanded into a multinational research effort. From its inception, the Human Genome Project included the genomes of five model organisms: the mouse, *Drosophila melanogaster*, *C. elegans*, *S. cerevisiae*, and *Escherichia coli*. These models were included to help develop and test the procedures needed for the much more complex human genome. The genomes of *C. elegans* and *Drosophila* were completed in 1998 and 2000, respectively, and the mouse sequence is virtually completed. Since the Human Genome Project's inception, many additional organisms have been selected for sequencing.

In addition to producing genomic sequence for these selected models, projects also produced related resources, including cDNA and genomic libraries, expressed sequence tags (ESTs), physical and genetic maps, stock centers, and databases. These resources engendered the enormous explosion in our understanding of the genetic regulation of cellular and biological processes. A critical component of these projects has been the open exchange of data and stocks. In fact, the NIH and the DoE established guidelines stipulating how and how frequently the sequence data were to be deposited in public databases. These databases enabled all investigators to access and manipulate these data. In addition to benefiting investigators who worked on a given model, the databases became major tools for all researchers. These genetic and genomic resources changed the way we identify new genes and predict their functions. Additionally, they promoted a new attitude of sharing and interactions among members of an organism's community and between members of different communities.

*Correspondence to: Steven L. Klein, Developmental Biology, Genetics and Teratology Branch, National Institute of Child Health and Human Development, Room 4B01, 6100 Executive Boulevard, Rockville, MD 20852. E-mail: kleins@exchange.nih.gov
Received 5 August 2002; Accepted 29 August 2002
DOI 10.1002/dvdy.10174

In 1997, the National Cancer Institute (NCI) of the NIH held a meeting of biomedical researchers to develop strategies for preclinical models for cancer research. One of the meeting's working groups discussed which nonmammalian models would provide insights into cancer research and which genetic and genomic tools would enable these models to do so. That group recommended that the NCI support research in five nonmammalian models: yeast, *C. elegans*, *Drosophila*, *zebrafish*, and *Xenopus*. They recommended that the NIH support the development of specific tools and resources for each of these models. For example, they recommended the creation of a *Xenopus* EST database, the development of improved husbandry to enable the production of biochemical quantities of eggs, and maintenance of inbred and genetically selected animals (http://www.nih.gov/science/models/nmm/nci_nmm_report.html). The recommendations were presented to the NIH director, who suggested that these tools and resources should not be limited to cancer research because their benefit could be much broader. Accordingly, he proposed a large meeting to enable a broad representation of the biomedical research community to recommend the tools and resources to enable nonmammalian models to provide new insights into all areas of biomedical research.

In 1999, the NIH director convened the Non-Mammalian Models (NMM) Meeting to evaluate the current status of genomic resources for the nonmammalian model organisms already undergoing genomic analysis, to identify additional resource needs for those organisms, and to consider what additional model organisms might be suitable for similar developments. At that meeting, approximately 80 leaders outlined broad goals for selecting animal models. They also devised a list of the genetic and genomic needs for each of five models (yeast, worm, fly, zebrafish, and *Xenopus*). These models were highlighted because of "their phylogeny, their experimental history, the size of their investigator community and the magnitude of their contribution to the sum of our biomedical knowledge." Members of each organism community gathered separately as a "Breakout Group" to develop a list of the tools and resources that were needed by their research community. Those recommendations were presented to the main meeting, developed into a report, and subsequently presented to the directors of all of the NIH institutes (http://www.nih.gov/science/models/nmm/index.html).

The *Xenopus* Breakout Group consisted of 10 *Xenopus* experts and of leaders of the mouse and zebrafish communities, who served as advisors. This group had formally solicited input from the *Xenopus* community for several months before the meeting. Based on the community's input, they made seven initial recommendations. They recommended the creation of an EST database, full-length cDNA sequences, microarrays, a database, genomic libraries, genomic resources for *X. tropicalis*, and training and stock centers (see http://www.nih.gov/science/models/nmm/appb6.html).

The *Xenopus* Breakout Group also considered the appropriate role for *X. laevis* and *X. tropicalis*. They concluded that *X. laevis* should be the main focus of the tools and resources. Their conclusion is reflected in the fact that they recommended 500,000 ESTs for *X. laevis* and only 50,000 for *X. tropicalis*, and full-length cDNA sequences only for *X. laevis*. However, they also concluded that *X. tropicalis* offered enormous potential as a genetic system. Accordingly, they also recommended the development of *X. tropicalis* genomic libraries and pilot genetic studies. Because these two strains have a high degree of sequence similarity and functional interchangeability, it was thought that resources for either would benefit both strains.

After the NMM meeting, the NIH took several concrete steps to support the recommended activities. For example, the NIH established a committee composed of representatives of each of the NIH institutes, the NMM committee, to promote interactions among NIH components and between the NIH and other agencies. It coordinates the recommended activities, considers new recommendations for those organisms that were represented at the NMM meeting, and considers including new organisms. Additionally, the NMM committee developed procedures for considering large grant application to produce genetic and genomic resources for NMM organisms (http://www.nih.gov/science/models/process/index.html). The NIH also established a working group for each of the model organisms. The National Institute of Child Health and Human Development (NICHD) formed the Trans-NIH *Xenopus* Working Group (XWG) because *Xenopus* is a major model for developmental biology research. The XWG consists of 11 NIH institutes who have a long history of supporting *Xenopus* research: NICHD; NCI; National Institute of General Medical Science (NIGMS); National Institute of Neurological Disorders and Stroke (NINDS); National Institute of Environmental Health Sciences (NIEHS); National Institute of Mental Health (NIMH); National Institute of Dental and Craniofacial Research (NIDCR); National Institute of Diabetes and Digestive and Kidney Disease (NIDDK); National Eye Institute (NEI); National Heart, Lung and Blood Institute (NHLBI); and the National Center for Research Resources (NCRR). The XWG enables the NIH to coordinate support for large *Xenopus* projects that are beyond the scope of a single institute. The XWG implements and oversees the activities recommended by the *Xenopus* community.

The initial recommendations of the NMM meeting enabled the XWG to implement the highest priority item, producing ESTs. However, before proceeding with the other recommendations, we needed to ensure that they reflected the needs of the broader *Xenopus* community. Additionally, most of the initial recommendations needed more detail. Accordingly, NICHD organized a large meeting of the *Xenopus* community

TABLE 1. Recommendations for *Xenopus* Research[a]

Genomic Tools	Development of the *X. tropicalis* System	Resource Centers
1) Produce ESTs From existing cDNA libraries and from new libraries prepared specifically for this purpose.	I. *X. tropicalis* Genetics Establish viability as a genetic system Insertional mutagenesis Chemical mutagenesis Deletions	A. Develop and maintain a database
2) Sequence full-length cDNA Inserts	II. *X. tropicalis* gene mapping PAC/BAC 8× coverage include fingerprint mapping RH panels; microsatellite markers; polymorphic markers	B. Animal stock center Maintain and distribute about 200 transgenic and mutant lines of *X. laevis* & *X. tropicalis;* include a training component
3) Microarrays a) Establish microarray facilities; these facilities should also distribute clones a) Provide supplements to existing grants for array readers	III. Supplement existing *X. laevis* grants to enable inclusion of and/or transition to *X. tropicalis* IV. Sequence the genome of *X. tropicalis*	C. Training center i. Continue and enlarge existing CSH lab *Xenopus* course ii. Establish new training facilities that emphasize techniques such as transgenesis

[a] The recommendations for *Xenopus* research developed by the participants of the 2000 NICHD *Xenopus* meeting. To ensure that none of the recommended activities would be overlooked, they were divided into three categories: Genomic Tools, Development of the *X. tropicalis* system, and Resource Centers. Distributing the recommendations into these categories, although somewhat inaccurate, emphasizes the importance of each activity. The complete recommendations and the background for each are described at http://www.nih.gov/science/models/Xenopus/reports/Xenopus_report.pdf. NICHD, National Institute of Child Health and Human Development; EST, expressed sequence tag; RH, radiation hybrid; CSH, Cold Spring Harbor.

the following year to further define the community's needs. Seventy representatives of the *Xenopus* community participated. The meeting had a session on each of the recommendations from the NMM meeting, namely ESTs, full-length cDNAs, microarrays, a database, needs for *X. tropicalis*, and training and stock centers. The participants in each session developed more detailed recommendations for those topics.

That meeting produced a report based around 10 detailed recommendations. Each recommendation was associated with one or more community contacts who offered to monitor their activity's progress, help the community revise the activity's goals as it proceeded and as other opportunities emerge, and communicate the community's evolving needs to the NIH. This report was distributed to the directors of each of the NIH's institutes. It has been instrumental in enabling the XWG to develop and support genetic and genomic resources for *Xenopus* research. The report is available on the NIH *Xenopus* Web site (http://www.nih.gov/science/models/*Xenopus*/reports/*Xenopus*_report.pdf).

PROJECTS

The report of the 2000 NICHD *Xenopus* meeting recommended that the NIH develop and support 10 activities for *Xenopus* research (Table 1). The XWG has been implementing those activities. To accomplish these goals, we interact extensively with the NCI, the National Human Genome Research Institute (NHGRI), the National Center for Biotechnology Information (NCBI), the Lawrence Livermore National Laboratory (LLNL), the DoE's Joint Genome Institute (JGI), and the Washington University Genome Sequencing Center. Below, we describe the status of these activities and explain how to access the information and resources that they are producing.

cDNA Libraries and ESTs

The community's highest priority was to develop a database of *Xenopus* ESTs to facilitate gene discovery, oligonucleotide-based knockout studies, analysis of normal and perturbed development, and mapping studies. They recommended ESTs for a series of embryonic stages and for numerous adult tissues. The community knew that many of the necessary cDNA libraries already existed in individual laboratories. Accordingly, they recommended collecting appropriate libraries from the labs that made them, making new libraries for necessary missing stages and organs, and using them to sequence ESTs.

To accomplish this goal, we arranged to use the infrastructure developed by the NCI to perform similar goals for two large projects, the Cancer Genome Anatomy Project (http://cgap.nci.nih.gov/) and the Mammalian Gene Collection (http://mgc.nci.nih.gov/index_html). This approach enabled us to begin very quickly and enables us to take advantage of the expertise of numerous on-going contracts. For example, *Xenopus* EST production can be continually optimized because NCI routinely identifies new contractors that offer higher throughput and lower cost.

Thirty-one cDNA libraries were donated by *Xenopus* labs, and an additional 17 libraries were constructed by Life Technologies from tissues donated by *Xenopus* labs. These 48 libraries represent developmental stages from oocyte to stage 32, and eight adult organs;

TABLE 2. IMAGE *Xenopus* cDNA Libraries[a]

- *X. laevis*
  - Oocyte
    - NICHD_XGC_OOI
    - Wellcome/CRC pRN3 oocyte
    - *X. laevis* oocytes, stages 5–6 mixed
  - Egg
    - RIKEN *Xenopus* egg
    - Soares NXEG [Normalized]
    - Wellcome/CRC pcDNAI egg
    - Wellcome/CRC pSK egg
    - *X. laevis* unfertilized egg
  - Whole embryo
    - NICHD_XGC_Emb1 [Stage 10]
    - Kirschner embryo st. 10–14
    - Wellcome/CRC pSK st. 10.5
    - Wellcome/CRC pcDNAI st. 10.5
    - Wellcome/CRC pRN3 st. 10.5
    - X. laevis gastrula, stages 10.5–11.5
    - Wellcome/CRC pcDNAI st. 12
    - Xenla_13
    - Wellcome/CRC pRN3 st. 13–17
    - NICHD_XGC_Emb2 [St. 17/18]
    - Harland stage 19–23
    - Wellcome/CRC pCS2+ st. 19–26
    - Wellcome/CRC pRN3 st. 19–26
    - NICHD_XGC_Emb3 [St. 24/25]
    - Wellcome/CRC pcDNAI st. 24–26
    - Wellcome/CRC pRN3 st. 24–26
    - NICHD_XGC_Emb4 [St. 31/32]
  - Adult tissues
    - NICHD_XGC_Bm1 [brain]
    - NICHD_XGC_Eye1 [eye]
    - NICHD_XGC_He1 [heart]
    - NICHD_XGC_Kid1 [kidney]
    - NICHD_XGC_Li1 [Liver]
    - NICHD_XGC_Lul [Lung]
    - Harland ovary
    - NICHD XGC Ovl [ovary]
    - NICHD_XGC_Spl [spleen]
  - Hyperdorsalized embryos
    - Xenla_13LiCl [St. 13, LiCl-treated]
    - Cho Li-treated gastrula
  - Embryo regions
    - Wellcome/CRC pSK animal cap
    - Wellcome/CRC pRN3 dorsal lip
    - Wellcome/CRC pRN3 head (st. 30)
    - Wellcome/CRC pRN3 tail (st. 30)
  - Subtracted libraries
    - Wellcome/CRC pRN3 st. 13–17 egg/animal cap [egg library minus st. 13–17 An. Cap]
    - Wellcome/CRC pRN3 st. 19–26 egg/animal cap [egg library minus st. 19–26 An. Cap]
- *X. tropicalis*
  - Egg
    - Wellcome/CRC pCS107 tropicalis egg
  - Embryo
    - Wellcome/CRC pCS107 tropicalis st. 10–12
    - NICHD_XGC_Emb5 [St. 10–13]
    - NICHD_XGC_Emb6 [St. 14–19]
    - NICHD_XGC_Emb7 [St. 20–27]
    - NICHD_XGC_Emb8 [St. 40–45]

[a]The official names of the *Xenopus* libraries available to the community through the IMAGE Consortium, listed by embryonic stage and tissue type. In most cases, the library's name indicates its provider and source. Libraries designated "NICHD_XGC_. . . " are newly constructed. Additional details are shown as needed, in brackets. A complete description of each library is at http://image.llnl.gov/image/html/*Xenopus*lib_info.shtml.

42 of the libraries are from *X. laevis* and six are from *X. tropicalis* (Table 2, and http://image.llnl.gov/image/html/Xenopuslib_info.shtml). The libraries were arrayed at LLNL, made part of the IMAGE consortium (http://image.llnl.gov/), and are available from Research Genetics (http://www.resgen.com/intro/clones.php3). Libraries from some of the originally recommended stages (tadpole), organs (pancreas, fat body), and cell lines (XTC) have not been made yet; they will be constructed soon.

All of these cDNA libraries are being used to produce ESTs. The first 115,000 ESTs were sequenced at the Washington University Genome Sequencing Center (http://genome.wustl.edu/est/index.php?xenopus=1). Currently, ESTs are being sequenced at the NIH Intramural Sequencing Center (NISC; http://www.nisc.nih.gov/index.html) and at Agencourt Biosciences Corporation (http://www.agencourt.com). All of these ESTs are submitted to dbEST, which currently lists over 215,000 *X. laevis* and over 160,000 *X. tropicalis* ESTs (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). The original recommendation was to produce 550,000 ESTs. This project is ongoing.

When the number of *X. laevis* ESTs in dbEST reached 100,000 in early 2001, we asked members of the community if they wished to reprioritize the recommendations or had suggestions for additional resources. They recommended that we continue sequencing ESTs from the existing libraries, make libraries from more of the originally recommended stages and adult tissues, and also construct a UniGene set. Accordingly, we continued with the original EST plan and asked NCBI to construct a UniGene set for *Xenopus*. NCBI constructed the set from the 150,000 *X. laevis* sequences in dbEST and GenBank. The set is periodically up-dated (see http://www.ncbi.nlm.nih.gov/UniGene/Xl.Home.html). It currently consists of over 18,000 unique genes (Table 3). However, there are fewer than 2,000 genes with a characterized or full-length insert sequenced mRNA, emphasizing the need for full-length insert sequencing. Many genes are rarely expressed, so that many interesting transcripts will have to be identified from a single EST. Furthermore, many genes are represented numerous times. This level of redundancy means that care needs to be taken to ensure that the same genes are not sequenced repeatedly. A UniGene set will soon be constructed for *X. tropicalis*.

Mutagenesis and Phenotyping of X. tropicalis

The community's highest priority for *X. tropicalis* was establishing that it was amenable to genetic manipulations. To accomplish this goal, they recommended pilot projects performing mutagenesis and phenotyping. Ultimately, *Xenopus* lines with mutations in specific genes will enable detailed examination of gene regulation and interaction in embryonic patterning and organogenesis.

TABLE 3. Statistics for *X. laevis* UniGene[a]

A. Sequences included	
2,340	mRNAs
71,159	EST, 3′reads
87,217	EST, 5′reads
+ 1,553	EST, other/unknown
162,269	total sequences in clusters
B. Final number of clusters	
18,984	Clusters total
1,849	Clusters contain at least one mRNA
18,622	Clusters contain at least one EST
1,487	Clusters contain both mRNAs and ESTs
C. Histogram of cluster sizes	
Cluster size	Number of clusters
1	5276
2	2367
3–4	3336
5–8	3407
9–16	2426
17–32	1350
33–64	566
65–128	185
129–256	52
257–512	15
513–1024	2
1025–2048	1
2049–4096	1

[a] Known genes are from GenBank and ESTs are from dbEST (both, as of July 31, 2002; UniGene build #19). Section A shows the number of sequences in UniGene clusters (sequences from the same gene), by mRNA and EST sequences. There are approximately equal numbers of 3′ and 5′ sequences. But, only about 1% of the sequences contain a complete mRNA. Section B shows the number of UniGene clusters. There are about 19,000 unique genes in this set, and 98% contain at least one EST. GenBank contains fewer than 2,000 *X. laevis* genes with a characterized or a full-insert sequenced mRNA. Section C shows the number of distinct sequences (Cluster size) present in each cluster (Number of clusters). Twenty-eight percent of the clusters (5,276 clusters) have only one sequence (probably an EST, Section B), and over 250 genes are represented more than 65 times each. EST, expressed sequence tag.

To address this recommendation, the XWG issued a Request for Applications in 2001. It solicited applications to examine the feasibility of using *X. tropicalis* for standard genetic manipulations. Specifically, it sought applications to optimize the conditions required to perform efficient, large-scale mutagenesis and to use the optimized parameters to perform small-scale mutagenesis; phenotyping; and gene cloning, identification, and characterization. The Request for Applications also required applicants to make their mutant animals, protocols, and data available to the community.

Three projects were funded recently. Collectively, they will optimize the parameters for making deletions, insertions, and point mutations with ENU. The deletion project will devise the procedures needed to produce deletion mutations throughout the genome with gamma rays. Deletion mutations are advantageous because they produce null phenotypes. These can be identified readily and used to discover the responsible genes. The mutants produced by this project will be a valuable resource for the community because they will help to map and characterize the *X. tropicalis* genome. The insertion project will optimize the parameters required to produce random insertional mutations with transposase- and retroviral-mediated integration systems. The major advantage of this strategy is that it enables rapid gene identification and cloning because the inserted gene is tagged with a traceable marker. This approach will enable identification of tissue-specific genes that play critical roles in organogenesis. The ENU project will devise the parameters required to use the powerful chemical mutagen, ENU. ENU is the most efficient means of producing point mutations throughout the genome. This strategy will enable the production of large numbers of mutant animals. Thus, these projects will play a significant role in devising the procedure required to perform mutagenesis in *Xenopus* and in generating mutants. They will share with the community their protocols, results, and mutant animals. Information about each project's progress and about how to obtain the resources that they are producing will be explained on the projects' Web sites, which are being set up now. Links to those Web sites will be available from the NIH Xenopus Web site (http://www.nih.gov/science/models/xenopus/).

Database

The highest priority under the "Resource" category was the production of a *Xenopus* database. Several components of the NIH currently support databases for model organisms. We are eager to receive applications to support a *Xenopus* database. Potential applicants should follow the NIH process for considering large applications to support model organism research (http://www.nih.gov/science/models/process/index.html).

Sequence Full-Length cDNA Inserts

The community recommended sequencing full-length cDNA inserts because they will be extremely useful for expression cloning. Specifically, they recommended producing full-length sequences of libraries from oocyte and egg, gastrula, neurula, tail bud, and tadpole stages.

We have begun this project by determining which UniGene and IMAGE *X. laevis* libraries (n = 39) contain full-length inserts by comparing their 5′ ESTs to reference sequences (characterized *X. laevis* mRNAs with a 5′ untranslated region longer than 100 nucleotides). Caveats with these analyses include the facts that there is a small number of *Xenopus* reference sequences and that several of the libraries have few ESTs. However, the results indicate that 80% of these libraries have a high probability of containing full-length inserts (i.e., have a full-length fraction of greater than 40%, which is the minimum required for acceptance into the MGC; see Table 4). Among these

TABLE 4. Details of Full-Length Fraction Analysis of *X. laevis* UniGene Libraries[a]

Seqs	%Divers	%CDS-OK	InLen	AlnLen	%FLDivers	Library
1128	76.7	0.917	2977.9	346.8	70.3339	NICHD_XGC_Ovl
17	64.7	1.000	1892.5	514.5	64.7	*Xenopus*_laevis_ZAP_Express_endodermal_cDNA_library
1430	64.0	1.000	1638.0	329.5	64	RIKEN_*Xenopus*_egg
3552	132.1	0.419	1209.8	358.0	55.3499	*Xenopus*_laevis_oocyte
458	69.4	0.737	1447.5	419.5	51.1478	Wellcome_CRC_pRN3_St10_5
63	60.3	0.750	1493.5	629.5	45.225	Harland_stage_19–23_*Xenopus*_laevis_cDNA
1731	68.4	0.639	1950.4	405.1	43.7076	NICHD_XGC_Emb2
1538	53.4	0.750	1786.2	395.1	40.05	Wellcome_CRC_pRN3_oocyte
460	95.7	0.412	1257.2	409.3	39.4284	*Xenopus*_neurula_plasmid_library
1917	51.1	0.737	1923.8	329.4	37.6607	*Xenopus*_EST_library
4697	53.7	0.662	1531.7	332.5	35.5494	NICHD_XGC_Spl
3527	54.8	0.643	2279.2	343.4	35.2364	NICHD_XGC_OOl
2389	50.8	0.673	1396.4	436.6	34.1884	Wellcome_CRC_pRN3_St19_26
2937	60.2	0.565	1829.4	339.9	34.013	NICHD_XGC_Emb1
5520	51.5	0.650	2100.6	338.9	33.475	NICHD_XGC_Emb4
2636	51.8	0.625	1363.6	414.7	32.375	Wellcome_CRC_pRN3_head
1711	53.9	0.574	1476.2	399.9	30.9386	Wellcome_CRC_pRN3_dorsal_lip
2807	53.9	0.558	1421.0	392.8	30.0762	Harland_stage_19–23
805	50.0	0.520	1465.4	431.6	26	Xenla_13LiCl
5452	41.4	0.615	1364.2	329.4	25.461	NICHD_XGC_Lu1
2709	51.2	0.485	1991.7	404.9	24.832	Wellcome_CRC_pSK_egg
7686	47.3	0.508	1638.0	471.0	24.0284	NIBB_Mochii_normalized_*Xenopus*_early_gastrula_library
736	45.6	0.526	1408.6	286.8	23.9856	Cho_Li_treated_gastrula
2862	55.9	0.429	1814.8	330.6	23.9811	NICHD_XGC_Emb3
2545	56.2	0.426	1766.3	359.4	23.9412	*Xenopus*_laevis_oocyte_non_normalized
2329	53.5	0.438	1255.8	399.9	23.433	Kirschner_embryo_St10_14
3551	35.8	0.644	1382.8	368.8	23.0552	NICHD_XGC_Li1
2892	54.0	0.413	1699.0	380.5	22.302	*Xenopus*_laevis_unfertilized_egg_cDNA_library
1985	38.8	0.561	1428.9	361.2	21.7668	Wellcome_CRC_pRN3_St13_17_egg_animal_cap
28718	35.7	0.551	1702.3	468.1	19.6707	NIBB_Mochii_normalized_*Xenopus*_neurula_library
3522	50.4	0.356	1130.5	333.1	17.9424	*Xenopus*_laevis_gastrula_non_normalized
469	41.9	0.368	1288.8	381.2	15.4192	Wellcome_CRC_pCS2+_st19–26
2383	53.3	0.281	1377.9	376.0	14.9773	Wellcome_CRC_pSK_animal_cap
1585	66.7	0.222	1004.1	308.6	14.8074	normalized_*Xenopus*_laevis_gastrula
35547	35.7	0.401	1642.3	490.3	14.3157	NIBB_Mochii_normalized_*Xenopus*_tailbud_library
3451	32.4	0.426	1519.5	308.2	13.8024	NICHD_XGC_He1
1023	54.1	0.250	1073.4	385.2	13.525	Wellcome_CRC_pSK_St_10_5
2277	47.0	0.255	1186.6	391.7	11.985	Wellcome_CRC_pcDNAI_egg
17678	28.6	0.142	1077.4	442.7	4.0612	Blackshear/Soares_normalized_*Xenopus*_egg_library

[a] UniGene libraries listed in order of the proportion of distinct sequences with full-length inserts (%FLDiver; column 6). The ratio of 'know-gene-complete ESTs'/'known-gene-complete ESTs plus known-gene-incomplete ESTs' (%CDS-OK; third column) is the fraction of full-length sequences in the library. Libraries with a full-length fraction of greater than 40% are worth further consideration. The number of distinct full-length clones expected per 100 ESTs (%FLDiver; sixth column) is the library's full-length fraction (%CDS-OK) times the library's diversity (%Divers). For libraries with a %CDS-OK greater than 40%, the %FLDiver figure provides that best indication that a library has many distinct full-length inserts. Definitions: Seqs, Sequences: Number of EST sequences; %Divers, Diversity: 100 × Genes/Clones; CDS-OK, Coding Sequence: Fraction of 5′ ESTs which are CDS-complete by comparison with Reference Sequences; InLen, Insert Length: Mean predicted insert length of those clones matching Reference Sequences; AlLen, Alignment Length: Observed alignment length of those clones matching Reference Sequences; %FLDivers, CDS-OK × %Divers: figure of merit for MGC libraries; EST, expressed sequence tag.

libraries are at least three from each of the recommended stages (except tadpole). An additional 10,000 ESTs are being sequenced from the best library of each of these stages to verify that they have full-length inserts. When this additional sequencing is complete, we will select candidate full-ORF clones from these libraries for full-insert sequencing. Additionally, we will consider sequencing full-length inserts from cDNA libraries specifically made to have full-length inserts.

When the *X. tropicalis* UniGene set is constructed, we will analyze the *X. tropicalis* libraries to indicate which contain full-length inserts. The best of those libraries will also be used for full insert sequencing.

Genomic Libraries

The community recommended the creation of *X. tropicalis* large insert genomic libraries. These are critical for cloning and analyzing genomic sequences, such as promoters, for clarifying the organization of groups of genes, and for mapping and cloning mutations. Additionally, these libraries are required for genome sequencing. We arranged to construct BAC libraries because their clones are relatively large and appear to represent faithfully an organism's genome.

To accomplish this goal, we arranged for *X. tropicalis* to be included in NHGRI's BAC Production Initiative

(http://www.nhgri.nih.gov/DER/BAC/BACResourceNetwork.html). A BAC library with 75-kb inserts was constructed by the Institute for Systems Biology from DNA of several 5th generation Nigerian *X. tropicalis* (male and female). Additionally, two BAC libraries with 150-kb inserts are being constructed with different restriction enzymes (*Eco*RI and *Mbo*I) by the Virginia Mason Research Center (http://vmresearch.org/lab_research/amemiya/libraries.htm). These large-insert libraries are being made from DNA of a single sixth generation inbred Nigerian *X. tropicalis*, who is a sister of the individual whose DNA is being used for whole genome shotgun sequencing (see below). These libraries will be made available for distribution to the community by the CHORI BAC/PAC distribution facility (http://www.chori.org/bacpac/home.htm).

Additionally, we have arranged for the Washington University Genome Sequencing Center to produce Fingerprint Contig maps (fpc) and BAC end sequences (bes) of these libraries. They will fpc map and end sequence approximately 150,000 clones from the large-insert libraries and end sequence approximately 75,000 clones from the small-insert library. These fpc and bes data will be made available and the BAC map will be displayed by means of a distributed annotation server. Sequence traces will also be deposited into NCBI's Trace Archive. Additionally, these end sequences and pfc will be used to sequence the *X. tropicalis* genome, as described below.

Radiation Hybrid Panels and genetic maps of *X. tropicalis* were also recommended. We are eager to consider applications to produce these.

Stock Center

The National Center for Research Resources (NCRR) supports stock centers for many organisms. They are interested in considering applications to develop a *Xenopus* Stock Center. Potential applicants should follow the NIH process for considering large applications for model organisms (see http://www.nih.gov/science/models/process/index.html).

Microarrays

The recommendations included establishing microarray facilities and providing supplements to existing grants to enable them to purchase array readers. Several components of the NIH currently support microarray facilities for model organisms, and we are eager to consider application to support *Xenopus* microarray facilities. Potential applicants for a microarray facility should follow the NIH process for considering large applications for model organisms (see http://www.nih.gov/science/models/process/index.html). Additionally, the institutes of the XWG will consider providing supplements to existing grants that they support to enable the addition of microarray readers when they are to be used to accomplish the goals of the original project.

X. tropicalis Transition Supplements

The community recommended that the NIH make supplemental funds available to NIH-funded *X. laevis* labs to enable them to include *X. tropicalis* in their studies. It was recommended that modest additional funds would enable a *X. laevis* lab to include *X. tropicalis*. Accordingly, we have announced the availability of these funds, for example at international meetings. There have been very few requests for these supplemental funds. We continue to be interested in considering requests for these supplements to existing NIH grants.

Training Center

The community recommended continuing the Cold Spring Harbor Lab's *Xenopus* course, which we have supported for the past 10 years, to enlarge the course, and to support additional courses that emphasize techniques, such as transgenesis. We are eager to consider applications to provide this training.

Genome Sequencing

We began having conversations in 2000 with the JGI about sequencing the *X. tropicalis* genome. In 2001, the JGI had completed several genome projects, including *Fugu rubripes* as part of an international consortium. The JGI was interested in sequencing a large genome that would complement their Functional Genomics Program focused on gene regulatory networks. Through discussions with members of the *Xenopus* research community and the JGI Scientific Advisory Board, *X. tropicalis* was chosen for genome sequencing. The project was outlined and more fully developed at a meeting of *Xenopus* leaders in March of 2002 at the JGI Production Sequencing Facility in Walnut Creek, California. That meeting enabled the community to express their needs and expectations from a sequencing project and enabled the JGI to elaborate their interests and capabilities. Additionally, the meeting established an advisory board consisting of community representatives and JGI personnel to disseminate information to the community, to assess community awareness and expectations and to relay the community's input to the JGI. The goal of the *X. tropicalis* genome project is to produce high-quality sequence and annotation that will meet the needs of the research community.

JGI has begun this project by producing two sets of whole genome shotgun libraries: 3- to 4-kb and 8- to 10-kb inserts in plasmid vectors. These libraries were made from DNA of a single sixth generation inbred Nigerian *X. tropicalis*, who is a sister of the individual whose DNA is being used for the 150-kb BAC libraries. The DNA was isolated from several tissues, including red blood cells, muscle, and bone. JGI will sequence the whole genome libraries to 5-6× and will incorporate additional BAC end sequences into an assembly. Depending on the contiguity of the assembly, additional sequencing to a depth of 8× will be achieved with

additional shotgun coverage or more directed approaches by using BACs guided by the map and the assembly. The JGI has sequenced approximately 40 Mb of the small-insert (2- to 3-kb) library, and this sequence is being evaluated for quality control purposes. Initial results indicate a low level of mitochondria contamination and an overall GC content of 42%. Several repeat sequences have been identified, and a list of these is available at the JGI Web site (www.jgi.doe.gov). In addition, the JGI has selected 10 BACs from the 75-kb BAC library for shotgun sequencing and assembly. These BACs are currently being finished to produce a single contig for each insert. These sequences are available for blast searches on the JGI Web site (www.jgi.doe.gov) and reads are deposited monthly in the trace depository at NCBI. In addition, searchable browsers are available for each BAC and display tracks for homologies to human proteins, gene models, and ESTs.

CONCLUSIONS

Over the past few years, the NIH has had many discussions with the *Xenopus* community to determine which genetic and genomic resources and tools are needed for *Xenopus* research. In 2000, the community presented a list of 10 prioritized recommendations. Since then, the XWG has been initiating and supporting those recommended activities. We arranged for the collection and production of cDNA libraries, the production of ESTs, the production of genomic libraries, pilot genetic screens, and genome sequencing. Currently, approximately 50 cDNA libraries are available from the IMAGE consortium; over 375,000 ESTs are in dbEST; three genomic libraries are being constructed, will be fpc mapped and end sequenced, and made available; sequencing of full-length cDNA inserts is beginning; three pilot genetic screens have begun; and the sequencing of the *X. tropicalis* genome is beginning. We are interested in receiving applications to perform those activities that require investigator-initiated applications, namely a database, stock center, microarray facility, supplements, and training programs. Potential applicants should contact the communicating author.

We welcome an on-going dialogue with the community to continually reassess needs and priorities. Are the original recommendations still appropriate or should they be modified, re-prioritized, or augmented? For example, the current recommendations put a greater emphasis on resources for *X. laevis* than on resources for *X. tropicalis*. Is that still the appropriate distribution of resources? Additionally, this dialogue must reflect the needs of all components of the community and not of only a few areas.

The resources that have been developed for *Xenopus* are enabling new analyses of *Xenopus* genes and gene function. They are also engendering waves of new projects examining *Xenopus* genetics and genomics. These new resources and projects have led to a new level of activity and excitement in *Xenopus* research. They are bringing geneticists and genome scientists into the *Xenopus* field and are re-emphasizing the advantages of *Xenopus* to other research communities. Collectively, these activities enable *Xenopus* to contribute to genetics research as it does to cell and developmental biology research. The ability of the *Xenopus* model to make significant contributions to these three fields emphasizes its unique and important role in biomedical research.